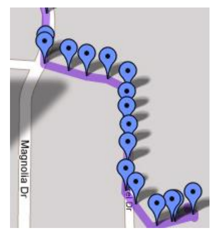
National Center for Transit Research

TRAC-IT:

Travel Behavior Data Mining Using GPS-Enabled Mobile Phones



Sean Barbeau

Research Associate

Center for Urban Transportation Research

University of South Florida

Opportunities

- Proliferation of cell phones
  - 61% of the world's population (4.1 billion) and 89% of U.S. (276.6 million) are mobile subscribers (Jun. 09) [1][2]
  - 23% of U.S. Households are Wireless–Only (Dec. 09) [3]
  - E-911 mandate for locating cell phones
- Proliferation of cell phone "apps"
  - While data is being collected from participant via phone, location-aware mobile apps can provide services to user (e.g. personalized traffic reports)
    - Incentive for extended survey participation
    - Longer survey period s with smaller samples for study

[1] International Telecommunications Union, "Measuring the Information Society - The ICT Development Index," International Telecommunications Union, 2009. [PDF]. Available: http://www.itu.int/ITU-D/ict/publications/idi/2009/material/IDI2009_w5.pdf. [
[2] http://www.ctia.org/media/industry_info/index.cfm/AID/10323
[3] http://news.yahoo.com/s/ap/20091216/ap_on_hi_te/us_cell_phones_only

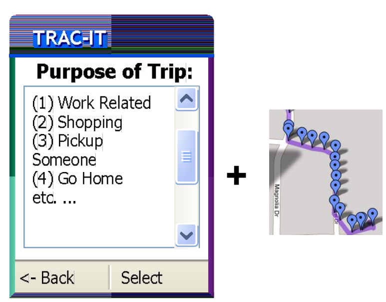
TRAC-IT

- Mobile software for GPS-enabled phones
  - Like an iPhone App
  - It's **<u>OPT-IN</u>**
- Features:
  - Runs on low to high tier phones
  - Records a person's travel behavior (an electronic activity diary)
  - Collects O/D and route information via GPS for **<u>all modes</u>**
  - Increases **<u>quality</u>** and **<u>quantity</u>** of collected information
  - Provides "hyper-personalized" real-time travel information services (e.g., traffic alerts)

USF
UNIVERSITY OF SOUTH FLORIDA

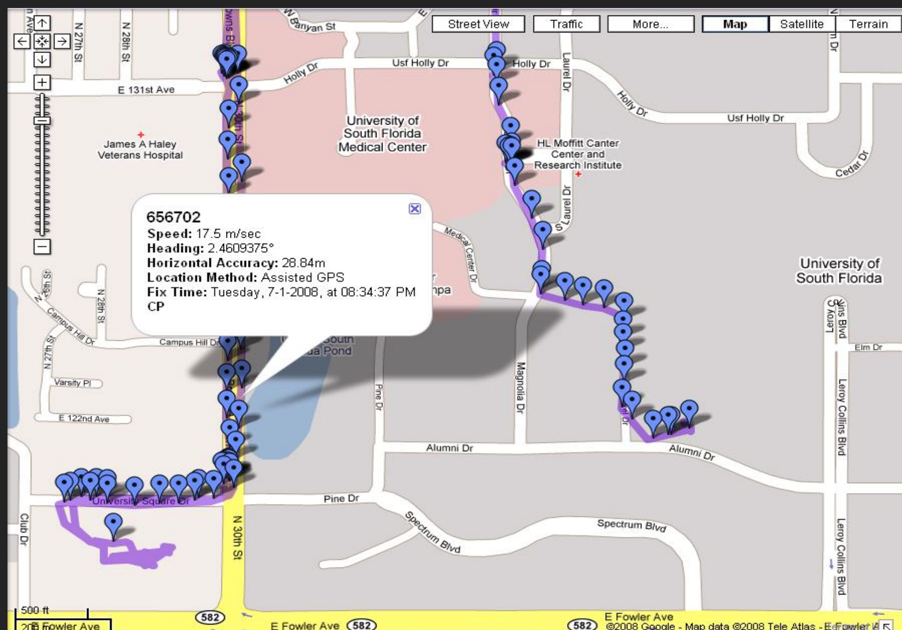
TRAC-IT

- Two modes for TRAC-IT:
  - **PASSIVE**
    - No interactions with user, runs in background
    - Records GPS path, provides real-time services
  - **ACTIVE**
    - Adds questions at the end of their trips:
      - Name for location
      - Mode of Transportation
      - Purpose of Trip
      - Occupancy of Vehicle



USF
UNIVERSITY OF SOUTH FLORIDA

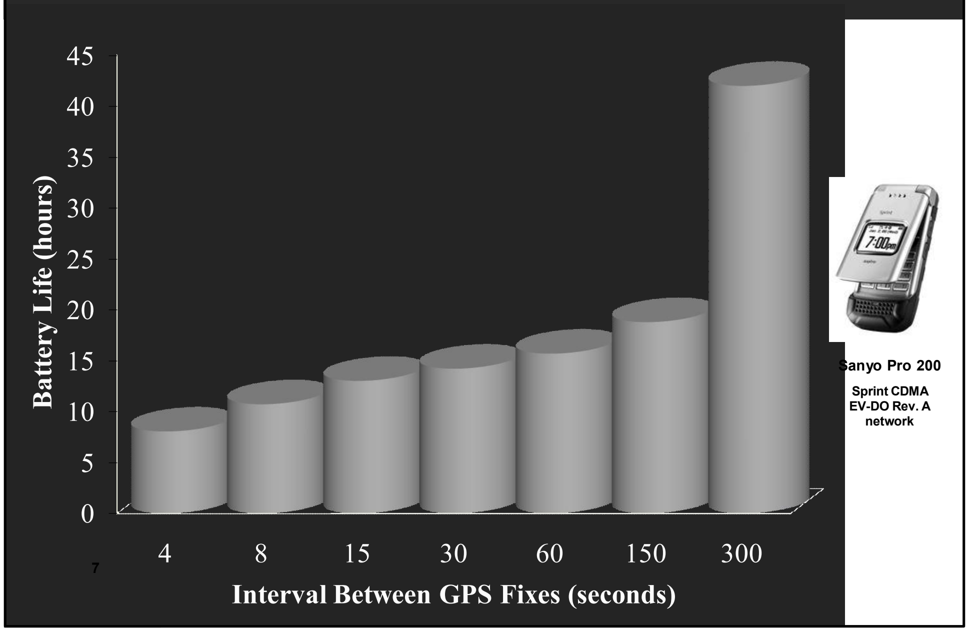
Assisted GPS data from TRAC-IT



GPS Data Pre-Processing

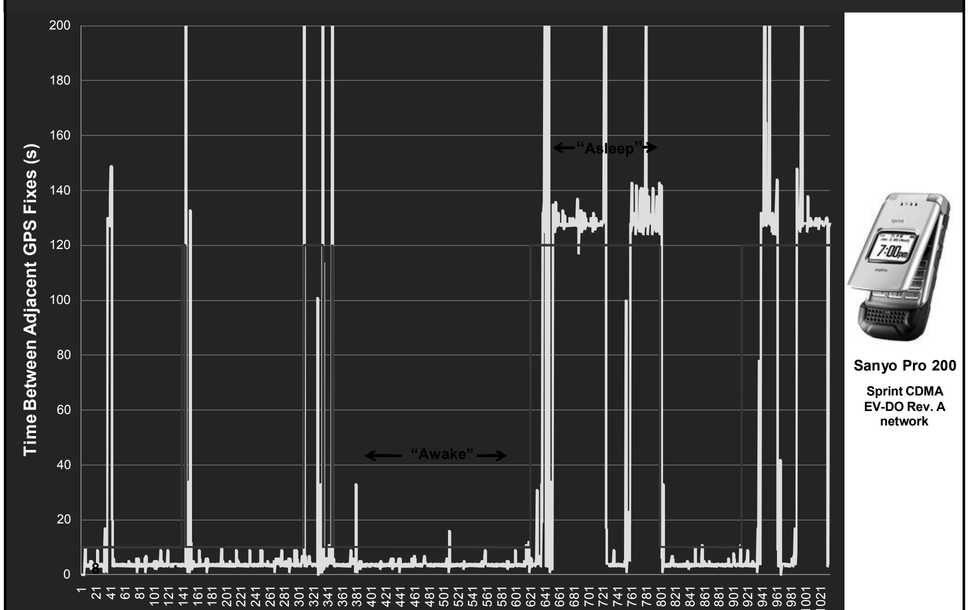
- Battery life is key concern for mobile apps
- If the user's phone dies, they will not use the app
- Problems with tracking:
  - GPS consumes significant energy for each fix
  - Wireless communication drains battery fast
- Solution:
  - Create data **pre-processing algorithms** that run **on** the cell phone **before** data is sent to server

Impact of GPS on Battery



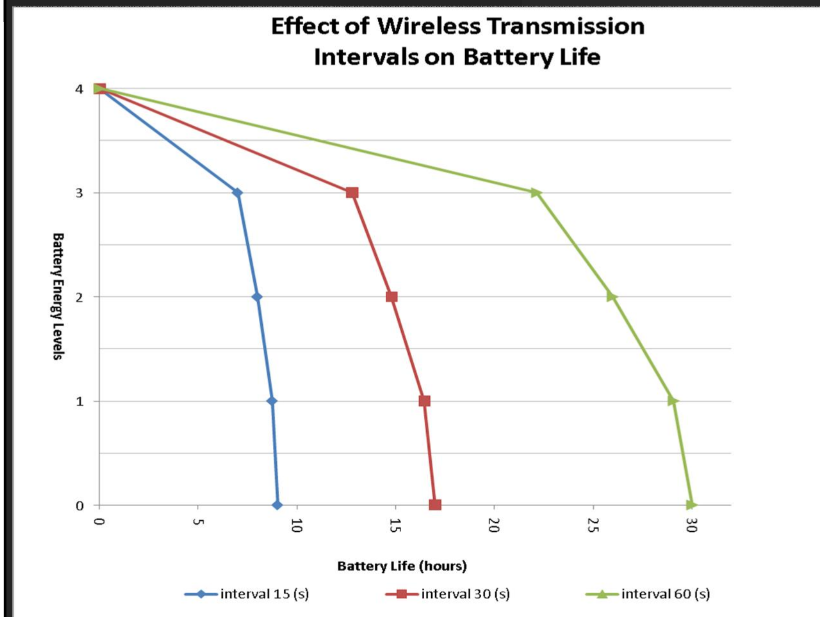
7

Solution: GPS Auto-Sleep



8

Impact of Wireless on Battery



Sanyo 7050

Sprint CDMA
1xRTT
Network
UDP

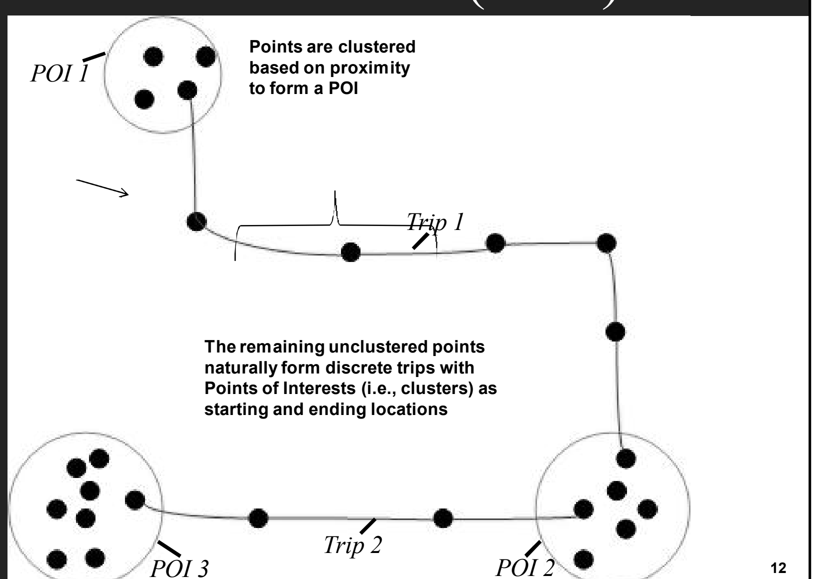
Solution: Critical Point Algorithm

10

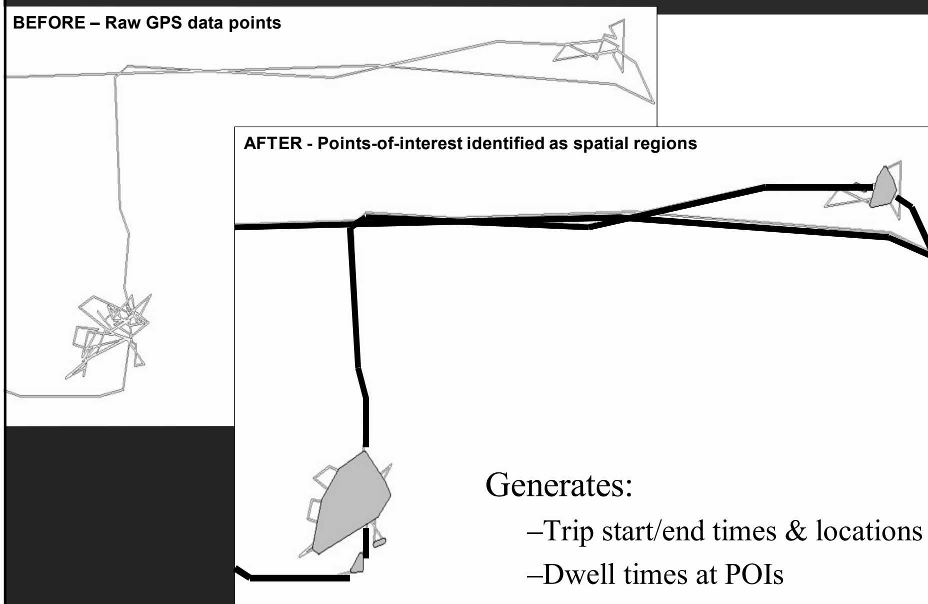
GPS Data Post-Processing

- Once the GPS data reaches the server, it is stored as records in a database
  - (X, Y) coordinates
- In order to derive information from GPS, **spatial data mining** is necessary
- **Automation** is key for large datasets!
- Algorithms based on spatial operations can use spatial databases (e.g., PostGIS)

Hierarchical Clustering can find Points-of-Interest (POIs)



Ex. Car Trip

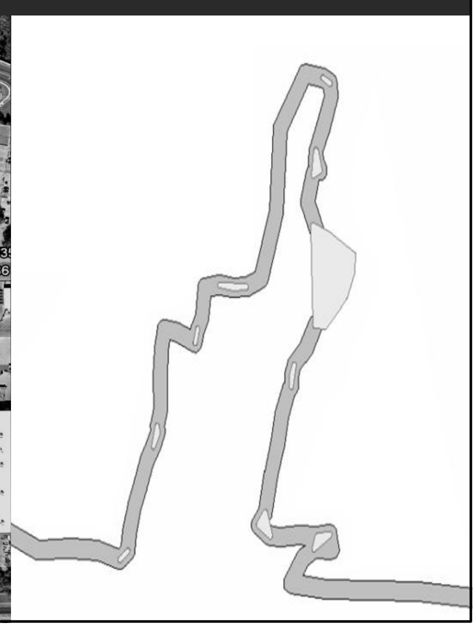


Generates:

- –Trip start/end times & locations
- –Dwell times at POIs

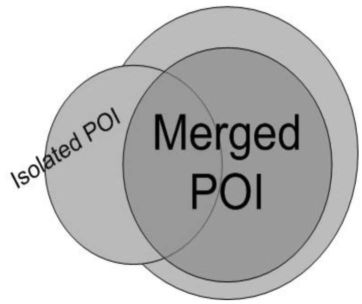
13

Ex. Walking trip



Merging User POIs

- Multiple visits to the same “location” should be registered with same POI
- Needed to count visitation frequency
- Algorithm uses POI overlap/bounding boxes to merge similar POIs
- Can be per user, or aggregate



UNIVERSITY OF SOUTH FLORIDA

15

Deriving Trip Characteristics from GPS data

- Passive tracking places least burden on participant
- However, surveyors often need additional data beyond GPS:
  – Mode of Transportation
  – Purpose
  – Occupancy
- Can we automatically determine these characteristics just from GPS?

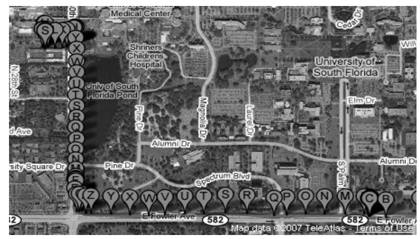
UNIVERSITY OF SOUTH FLORIDA

16

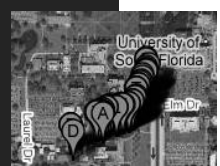
Automated Mode Detection

Can artificial neural networks identify MODE from GPS data alone?

Bus Trip



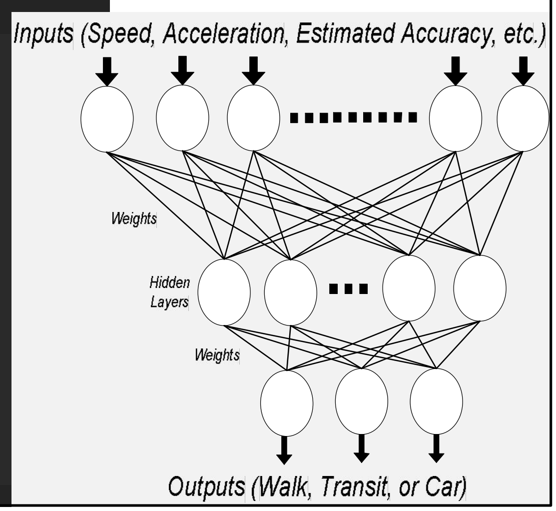
Car Trip



Walking Trip

Artificial Neural Networks

- AI tool for data-driven machine learning
- Two Step Process:
  - Training with known example data
  - Testing with new, unseen data

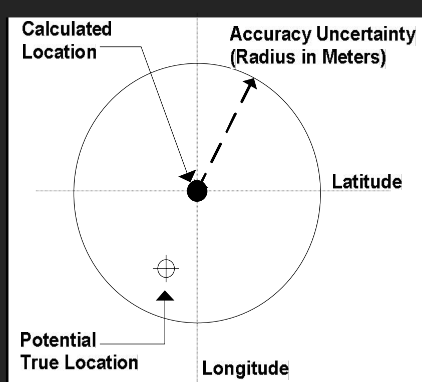


Sample Input Data

Latitude	Longitude	Speed (m/s)	Heading (0-359)	Date & Time	Est. Accuracy (m)	Location Method
27.94330215	-82.33336639	13	286.52	2008-05-22 08:29:14.837	10.45	A-GPS
27.94348907	-82.33384704	7.75	292.34	2008-05-22 08:29:17.293	21.03	A-GPS
27.94371986	-82.33440399	5	298.57	2008-05-22 08:29:23.301	50.49	A-GPS
28.05500030	-82.40055847	-	-	2008-05-22 08:29:26.529	-	Cell-ID

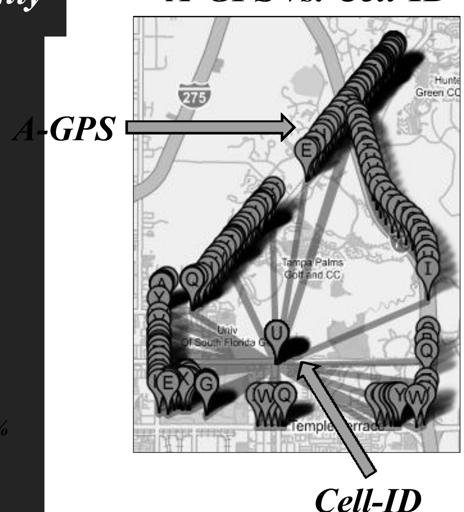
Input Data Attributes

Estimated Accuracy Uncertainty



***Probability of approximately 68%**

A-GPS vs. Cell-ID



20

Two Types of Datasets Studied

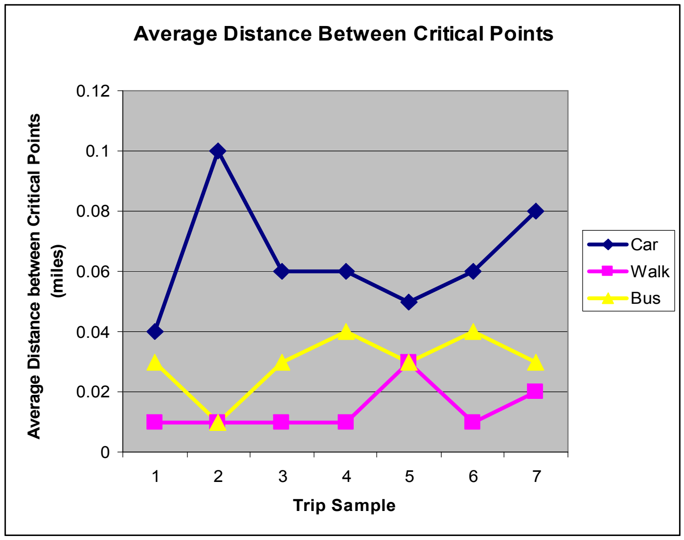
All GPS Points

Critical Points Only

Choosing Data Input Attributes

- User must choose data input attributes for neural network
- Goal is to find attributes that will easily identify modes of transportation
- Need to distinguish between similar modes
  - Especially Car vs. Bus

Choosing Data Input Attributes



Good for Car vs. Bus

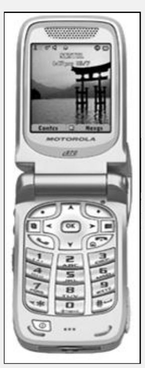
Final Inputs

- Inputs chosen for *All GPS Points* dataset:
  - *Avg. Speed*
  - *Max. Speed*
  - *Avg. Accuracy Uncertainty*
  - *Percent Cell-ID Fixes*
  - *Standard Deviation of Distance Between Stop Locations*
  - *Average Dwell Time*

Final Inputs

- Inputs chosen for *Critical Points Only* dataset:
  - *Avg. Speed*
  - *Max. Speed*
  - *Avg. Acceleration*
  - *Max. Acceleration*
  - *(# of Critical Points / Total distance of the trip)*
  - *(# of Critical Points / Total time of the trip)*
  - *Total Distance*
  - *Average Distance between critical points*

Experiment



- 114 trips recorded in Tampa, Fl
  - 38 car
  - 38 bus
  - 38 walk
- Devices = Motorola i870 and i580 phones
  - Sprint-Nextel iDEN network
- Software = TRAC-IT mobile app.
  - Java Micro Edition w/ JSR179 Location API
  - Queries position every 4 seconds

Experiment

- Neural Network Software = Weka
  – Used Java API for Multi-Layer Perceptron
- 10-fold cross validation used
  – Full data set randomly partitioned into 10 sets
  – 9 sets used for training, 1 set for testing
  – Repeated 10 times while alternating testing set
  – Reported accuracy is mean value of 10 tests

Results

- Numerous neural network settings were tested
- Best results:

Type of Input	*Accuracy*
All GPS Points	88.6%
Only Critical Points	91.23%

-Using .1 Learning Rate and 300 training epochs

Good for mobile phone battery!

Results

- **Breakdown of 91.23% accuracy for *Critical Point Only* dataset**

Mode of Transportation	***Average Accuracy Per Mode***
Car	92.11%
Bus	81.58%
Walk	100.0%

Similar traits

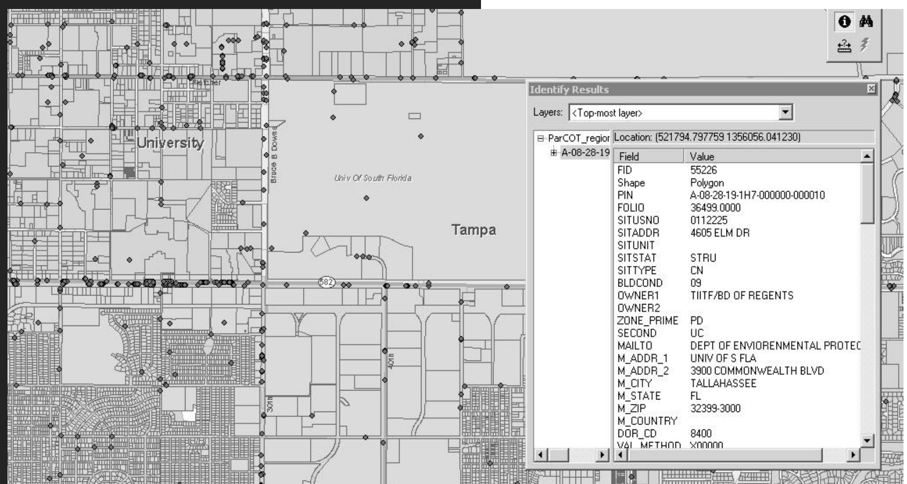
USF UNIVERSITY OF SOUTH FLORIDA

Automated Purpose Detection

- Use GIS Land-Use and Zoning maps to determine location type
  - Single-Family Home
  - Restaurant
  - Etc.
- Derive purpose from location type

USF UNIVERSITY OF SOUTH FLORIDA

Proof-of-concept Purpose Detection



- Used Hillsborough County Department of Revenue (DOR_CD) and Zone_Prime fields to programmatically identify land use

Sample DOR Codes

<u>**USE CODE**</u> <u>**PROPERTY TYPE**</u>

Residential

– 0000 Vacant Residential
– 0100 Single Family
– 0200 Mobile Homes
....

Commercial

– 1300 Department Stores
– 1400 Supermarkets
– 1600 Community Shopping Centers
– 1700 Office buildings, non-professional service buildings, one story
– 2000 Airports (private or commercial), bus terminals, marine, etc.
– 2100 Restaurants, cafeterias
– 2200 Drive-in Restaurants
2300 Financial institutions (banks, savings and loan companies, etc.)
– 2400 Insurance company offices
2500 Repair service shops (excluding automotive)
....

Institutional

7100 Churches
– 7200 Private schools and colleges
– 7300 Privately owned hospitals
– 7400 Homes for the aged
– 7500 Orphanages, other non-profit or charitable services
....

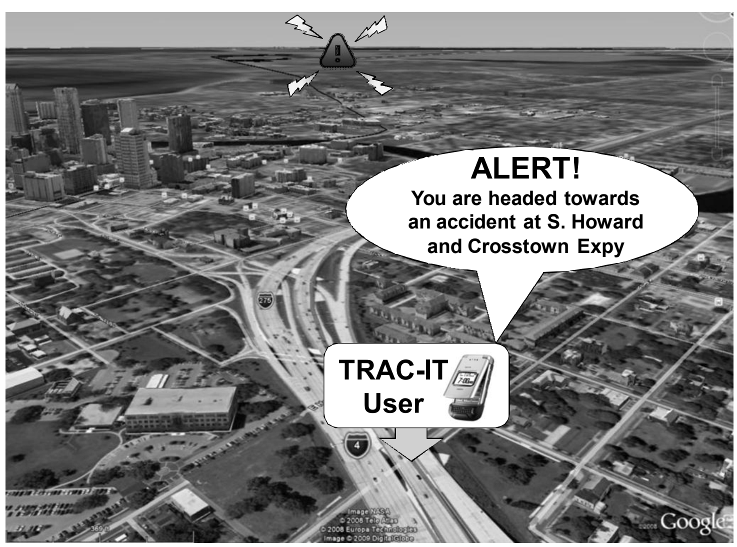
USF UNIVERSITY OF SOUTH FLORIDA

Automated Purpose Detection

- Possible, but many challenges:
  - Multi-use areas:
    - Baseball field at a school
  - Alternate uses:
    - Work at a restaurant
  - Coding issues:
    - Red Lobster designated as “Federal” instead of restaurant
- Likely useful for prompted recall
  - ~65% accurate in proof-of-concept

Provide Value to Participant

- **“Hyper-personalized” real-time traffic alerts**



34

What is Path Prediction?

- Real-time spatial data mining
- Predicts a user's real-time path using:
  - Real-time location
  - Historical travel behavior
- Based mainly on spatial data operations
- Once path is predicted, algorithm can find alerts along a traveler's predicted path
  - Ex. Traffic accidents, advertising
- Reduces irrelevant alerts sent to users

USF UNIVERSITY OF SOUTH FLORIDA

How It Works

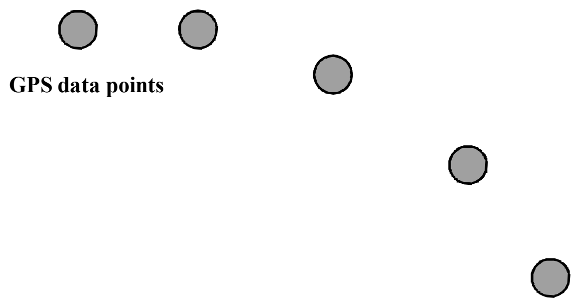
- Two steps:
  - **Part A** - Build user history over time from traveled paths
  - **Part B** – Predict immediate travel behavior based on real-time and historical travel behavior

USF UNIVERSITY OF SOUTH FLORIDA

© USF 2010
Patents Pending

How It Works (Part A)

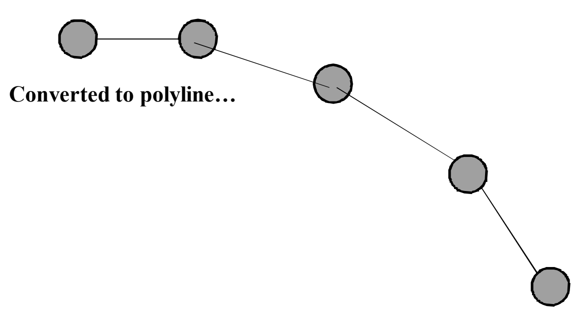
As user travels over time, recorded GPS data is translated into paths (polygons) in database



UNIVERSITY OF SOUTH FLORIDA

How It Works (Part A)

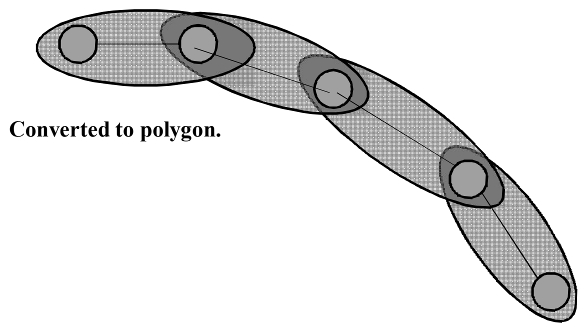
As user travels over time, recorded GPS data is translated into paths (polygons) in database



UNIVERSITY OF SOUTH FLORIDA

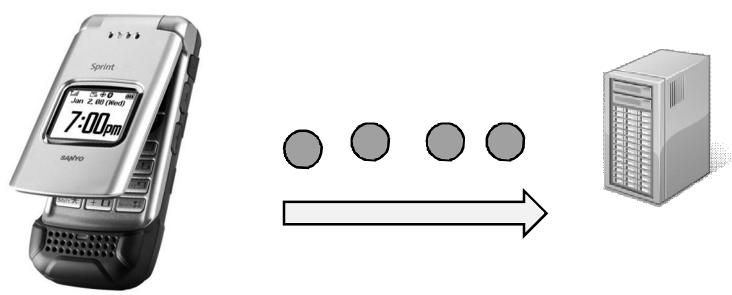
How It Works (Part A)

As user travels over time, recorded GPS data is translated into paths (polygons) in database



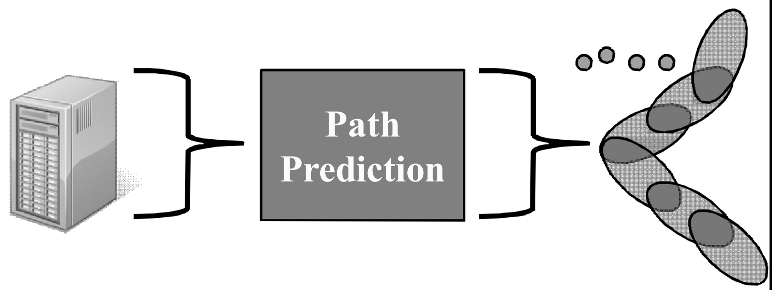
How It Works (Part B)

In real-time, phone sends GPS fixes to TRAC-IT server…



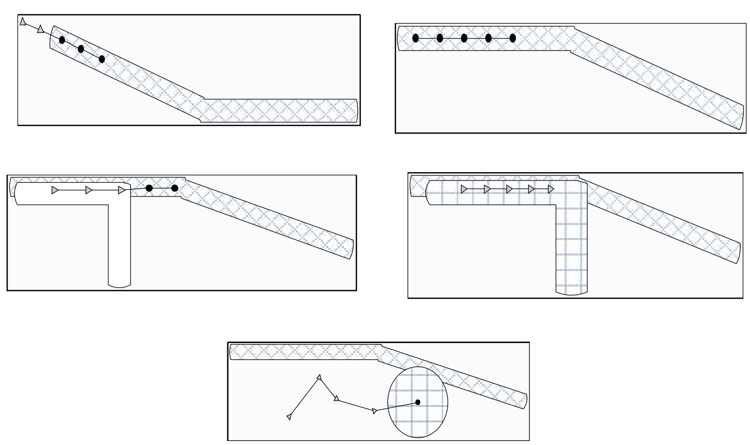
How It Works (Part B)

…Server runs Path Prediction and checks path history…



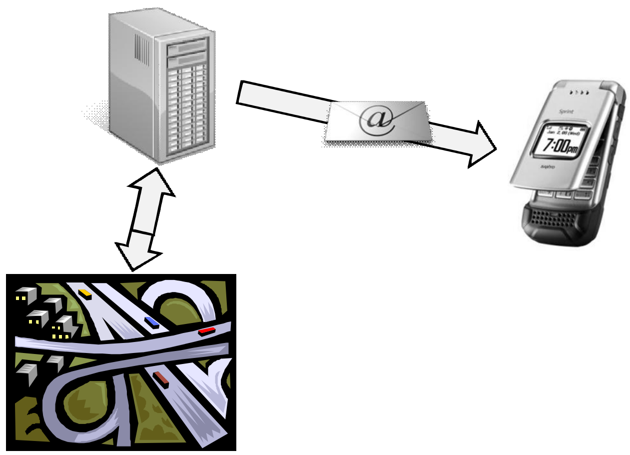
How It Works (Part B)

…An algorithm using a series of spatial operations identifies the most likely paths…



How It Works (Part B)

Server checks for incidents intersecting predicted paths from real-time data source, and alerts phone



© USF 2010
Patents Pending

43

© USF 2010
Patents Pending

44

UNIVERSITY OF SOUTH FLORIDA

UNIVERSITY OF SOUTH FLORIDA

Current Path Prediction Work

- Integrate Bayesian predictions based on POI visitation frequency with spatial predictions
- System integration with real-time travel information data sources in Florida

47

Acknowledgements

- Center for Urban Transportation Research (CUTR)
  - Phil Winters, Nevine Georggi
- USF Computer Science Department
  - Miguel Labrador, Rafael Perez
- National Center for Transit Research
- Florida Department of Transportation
- US Department of Transportation
- National Science Foundation
- Sprint-Nextel Application Developer Program

48

Questions?

Sean J. Barbeau, M.S.

Research Associate
Center for Urban Transportation Research
University of South Florida

(813) 974-7208

barbeau@cutr.usf.edu

USF Location-Aware Information Systems Lab:
http://www.locationaware.usf.edu/

USF UNIVERSITY OF SOUTH FLORIDA

For Additional Reading…

- Paola A. Gonzalez, Jeremy S. Weinstein, Sean J. Barbeau, Miguel A. Labrador, Philip L. Winters, Nevine L. Georggi, Rafael A. Perez. "Automating Mode Detection for Travel Behavior Analysis by Using GPS-enabled Mobile Phones and Neural Networks," *Institution of Engineering and Technology Intelligent Transportation Systems Journal*. doi: 10.1049/iet-its.2009.0029 *(to appear 2010).*
- Sean J. Barbeau, Nevine L. Georggi, Philip L. Winters. "TRAC-IT: Travel Behavior Data Collection using GPS-enabled Mobile Phones," *Human Factors 135 F – Quantifying Driving-Risk Exposure Committee Meeting at National Academy of Sciences' Transportation Research Board 89th Annual Meeting*. Washington, D.C., January 9th, 2010.
- Sean J. Barbeau, Miguel A. Labrador, Nevine L. Georggi, Philip L. Winters, Rafael A. Perez. "TRAC-IT: A Software Architecture Supporting Simultaneous Travel Behavior Data Collection and Real-Time Location-Based Services for GPS-Enabled Mobile Phones," *Proceedings of the National Academy of Sciences' Transportation Research Board 88th Annual Meeting,* Paper #09-3175. January, 2009.
- Narin Persad-Maharaj, Sean J. Barbeau, Miguel A. Labrador, Philip L. Winters, Rafael Perez, Nevine Labib Georggi. "Real-time Travel Path Prediction using GPS-enabled Mobile Phones," 15th World Congress on Intelligent Transportation Systems, New York, New York, November 16-20, 2008.
- Sean J. Barbeau, Miguel A. Labrador, Philip L. Winters, Rafael Perez, Nevine Labib Georggi. "Trac-It - A 'Smart' User Interface For A Real-Time, Location-Aware, Multimodal Transportation Survey," 15th World Congress on Intelligent Transportation Systems, New York, New York, November 16-20, 2008.
- Paola A. Gonzalez, Jeremy S. Weinstein, Sean J. Barbeau, Miguel A. Labrador, Philip L. Winters, Nevine Labib Georggi, Rafael Perez. "Automating Mode Detection Using Neural Networks and Assisted GPS Data Collected Using GPS-Enabled Mobile Phones, 15th World Congress on Intelligent Transportation Systems, New York, New York, November 16-20, 2008.
- Sean J. Barbeau, Miguel A. Labrador, Alfredo Perez, Philip Winters, Nevine Georggi, David Aguilar, Rafael Perez. "Dynamic Management of Real-Time Location Data on GPS-enabled Mobile Phones," Presented at UBICOMM 2008 – The Second International Conference on Mobile Ubiquitous Computing, Systems, Services, and Technologies, Valencia, Spain, September 29 – October 4, 2008. © 2008 IEEE.

http://www.locationaware.usf.edu/publications.htm

USF UNIVERSITY OF SOUTH FLORIDA